Table S1: CWT-1 mapping IGBP (rows) to generic PFTs (columns). Cells with the letter 'D' indicate a direct correspondence. Cells with LT indicate the need for leaf type information (broadleaf/needleleaf). Cells with 'C' indicate the need for C3/C4 partitioning. Several cells being filled means the IGBP class can map to fractions of generic PFTs.

IGBP to gPFT	TrBr Ev	TrBr De	TrNe Ev	TrNe De	ShBrEv	ShBr De	ShNe Ev	ShNe De	NatGr C3	NatGr C4	Crops C3	Crops C4
CRO											C	C
CSH					LT	LT	LT	LT	C	C		
CVM	LT	LT	LT	LT	LT	LT	LT	LT	C	C	C	C
DBF		D										
DNF				D								
EBF	D											
ENF			D									
GRA									C	C		
MF	LT	LT	LT	LT								
OSH					LT	LT	LT	LT	C	C		
SAV	LT	LT	LT	LT	LT	LT	LT	LT	LT	LT		
WSA	LT	LT	LT	LT	LT	LT	LT	LT	LT	LT		

Table S2: CWT-2 mapping generic PFTs (columns) to ORCHIDEE PFTs (PFT numbers), considering KG classes (rows).

	TrBrEv	TrBrDe	TrNeEv	TrNeDe	ShBrEv	ShBrDe	ShNeEv	ShNeDe	NatGr		Crops	
									Still-C3	Still-C4	LUH-C3	LUH-C4
KG Tropical	TroEBF	TroDBF	TemENF	TroDBF	TroEBF	TroDBF	TemENF	TroDBF	TroC3GRA	C4GRA	C3CRO	C4CRO
KG Arid Warm												
KG Arid Cool	TemEBF	TemDBF		TemDBF	TemEBF	TemDBF		TemDBF	TemC3GRA			
KG Temperate Warm												
KG Temperate Cool				BorDNF				BorDNF				
KG Boreal Warm	BorENF				BorENF				BorC3GRA			
KG Boreal Cool		BorDBF	BorENF			BorDBF	BorENF					

Table S3: CWT-3 KG 31 original classes (columns) to 7 aggregated classes (rows). The table indicates to which aggregated class a specific original class is entirely attributed, as specified by a '1' in the cell with the corresponding column and row.

KG31 to KG7	Af	Am	As	Aw	BSh	BSk	BWh	BWk	Cfa	Cfb	Cfc	Csa	Csb	Csc	Cwa	Cwb	Cwc	Dfa	Dfb	Dfc	Dfd	Dsa	Dsb	Dsc	Dsd	Dwa	Dwb	Dwc	Dwd	EF	ET
tropical	1	1	1	1																											
arid_warm					1		1																								
arid_cool						1		1																							
temp_warm									1			1			1																
temp_cool										1	1		1	1		1	1	1													
boreal_warm																			1			1	1			1	1				
boreal_cool																				1	1			1	1			1	1	1	1

Table S4: IGBP classification.

Code	Description

BSV	Barren Sparse Vegetation: Lands exposed soil, sand, or rocks and has less than 10% vegetative cover during any time of the year.
CRO	Croplands: Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.
CSH	Closed Shrublands: Lands with woody vegetation less than 2 meters tall and with shrub canopy cover >60%. The shrub foliage can be either evergreen or deciduous.
CVM	Cropland/Natural Vegetation Mosaics: Lands with a mosaic of croplands, forest, shrublands, and grasslands in which no one component comprises more than 60% of the landscape.
DBF	Deciduous Broadleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
DNF	Deciduous Needleleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
EBF	Evergreen Broadleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees and shrubs remain green year round. Canopy is never without green foliage.
ENF	Evergreen Needleleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.
GRA	Grasslands: Lands with herbaceous types of cover. Tree and shrub cover is less than 10%. Permanent wetlands lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.

MF	Mixed Forests: Lands dominated by trees with a percent cover >60% and height exceeding 2 meters. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60% of landscape.
OSH	Open Shrublands: Lands with woody vegetation less than 2 meters tall and with shrub canopy cover between 10-60%. The shrub foliage can be either evergreen or deciduous.
SAV	Savannas: Lands with herbaceous and other understory systems, and with forest canopy cover between 10-30%. The forest cover height exceeds 2 meters.
SNO	Snow and Ice: Lands under snow/ice cover most of the year.
URB	Urban and Built-Up Lands: Land covered by buildings and other man-made structures.
WAT	Water Bodies.
WET	Permanent Wetlands: Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water.
WSA	Woody Savannas: Lands with herbaceous and other understory systems, and with forest canopy cover between 30-60%. The forest cover height exceeds 2 meters.

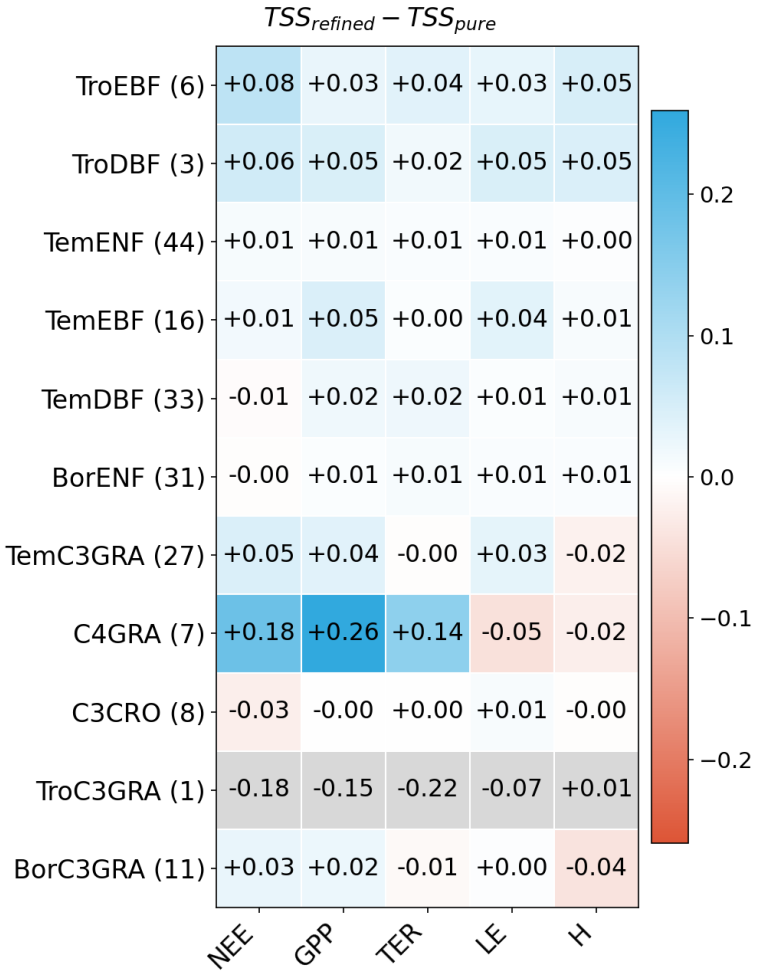


Figure S1: Difference of Taylor Skill Score (Taylor, 2001) between the refined and pure approach, per PFT (rows) and variable (columns). $\mathbf{TSS} = \frac{\mathbf{2(1+R)}}{(\boldsymbol{\sigma}* + \mathbf{1}/\boldsymbol{\sigma}*)^{\mathbf{2}}}$**, where** $\mathbf{R}$ **is the correlation between the observations and the model and** $\boldsymbol{\sigma}* = \frac{\boldsymbol{\sigma}_{\mathbf{model}}}{\boldsymbol{\sigma}_{\mathbf{obs}}}$ **is the ratio of their standard deviations (variability ratio). Positive values (in blue) indicate that the refined approach performs better.**

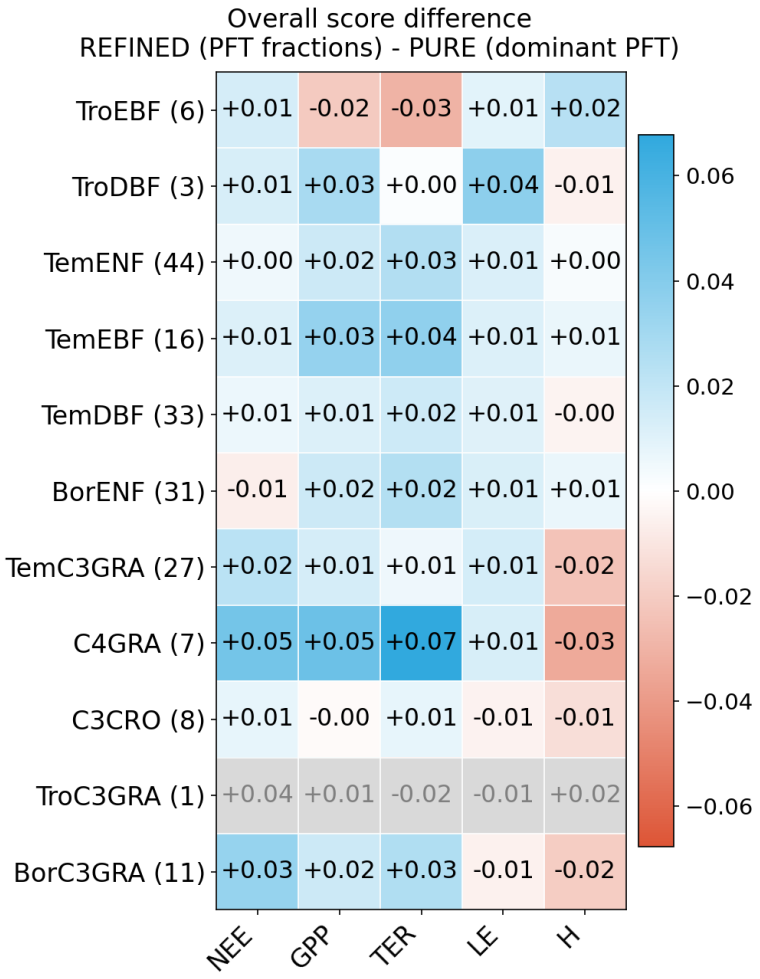


Figure S2: ILAMB-like overall score (Collier et al., 2018), per PFT (rows) and variable (columns), shown as the difference between the refined and the pure approach. The overall score is defined by combining 4 components score: $\mathbf{S_{overall}} = \frac{\mathbf{S_{bias}+2S_{rmse}+S_{phase}+S_{iav}}}{\mathbf{1+2+1+1}}$ **with** $\mathbf{S_{bias}} = \mathbf{exp}(-|\mathbf{bias}|/\boldsymbol{\sigma}_{\mathbf{ref}})$; $\mathbf{S_{rmse}} = \mathbf{exp}(-\mathbf{crmse}/\boldsymbol{\sigma}_{\mathbf{ref}})$, $\mathbf{S_{phase}} = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{1}+\mathbf{cos}\,(\frac{\mathbf{2\pi\Delta t}}{\mathbf{365}}))$ **and** $\mathbf{S_{iav}} = \mathbf{exp}(-\mathbf{iav_{norm}})$, **where** $\boldsymbol{\sigma}_{\mathbf{ref}}$ **is the standard deviation of the observations,** $\mathbf{crmse}$ **is the centralized RMSE,** $\boldsymbol{\Delta}\mathbf{t}$ **is the difference in days, between the timing of the maximum of the annual cycle in the model and in the reference product, and** $\mathbf{iav_{norm}} = |\mathbf{iav_{mod}} - \mathbf{iav_{ref}}| \,/\, \mathbf{iav_{ref}}$, **where** $\mathbf{iav_{mod}}$ **and** $\mathbf{iav_{ref}}$ **are the RMS of the model and reference anomalies after removal of the annual cycle. Positive values (in blue) indicate that the refined approach performs better.**